Mycologia, 102(4), 2010, pp. 887–897. DOI: 10.3852/09-198

Analysis of the complete mitochondrial genome sequences of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomiae*

Christine L. Stone

USDA-Agricultural Research Service, Foreign Disease-Weed Science Research Unit, 1301 Ditto Avenue, Fort Detrick, Maryland 21702

Martha Lucia Posada Buitrago[1]
Jeffrey L. Boore[2]

US DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, Walnut Creek, California 94598

Reid D. Frederick[3]

USDA-Agricultural Research Service, Foreign Disease-Weed Science Research Unit, 1301 Ditto Avenue, Fort Detrick, Maryland 21702

Abstract: The mitochondrial (mt) genomes of two soybean rust pathogens, *Phakopsora pachyrhizi* and *P. meibomiae*, have been sequenced. The mt genome of *P. pachyrhizi* is a circular 31 825-bp molecule with a mean GC content of 34.6%, while *P. meibomiae* possesses a 32 520-bp circular molecule with a mean GC content of 34.9%. Both mt genomes contain the genes encoding ATP synthase subunits 6, 8 and 9 (*atp6*, *atp8* and *atp9*), cytochrome oxidase subunits I, II and III (*cox1*, *cox2* and *cox3*), apocytochrome b (*cob*), reduced nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*), the large and small mt ribosomal RNA genes, one ORF coding for a ribosomal protein (*rps3*), and a set of 24 tRNA genes that recognize codons for all amino acids. The order of the protein-coding genes and tRNA is identical in the two *Phakopsora* species, and all genes are transcribed from the same DNA strand clockwise. Introns were identified in the *cox1*, *cob* and *rnl* genes of both species, with three of the introns having ORFs with motifs similar to the LAGLIDADG endonucleases of other fungi. Phylogenetic analysis of the 14 shared protein-coding genes agrees with commonly accepted fungal taxonomy.

Key words: *Basidiomycota*, codon usage, comparative genomics, gene order, genome organization, intron, LAGLIDADG motif, mitochondrial DNA

INTRODUCTION

Soybean rust represents one of the most significant threats to soybean production worldwide (Miles et al. 2003). Soybean rust was reported first in Japan in 1902, and the causal organism was identified as *Phakopsora pachyrhizi* Syd. and P. Syd. (Hennings 1903). The disease subsequently spread to other countries in southeastern Asia, Australia, Africa, South America and most recently to North America (Bromfield 1984; Levy 2005; Yorinori et al. 2005; Schneider et al. 2005).

Before the discoveries of *P. pachyrhizi* in South America and the United States, rust was detected on soybean in Puerto Rico in 1976 and identified as *P. pachyrhizi* (Vakili and Bromfield 1976). However isozyme studies (Bonde et al. 1988) suggested that there were two different species of *Phakopsora* causing rust on soybean, one in Asia and the eastern hemisphere and a second in the New World. Subsequent morphological characterization of the telia revealed that *P. pachyrhizi* was the cause of soybean rust in Asia, while the rust on soybean and other legumes in the New World was identified as a new species, *P. meibomiae* (Arthur) Arthur (Ono et al. 1992).

Recently these two *Phakopsora* species have been differentiated with species-specific PCR primers (Frederick et al. 2002). Sequence analysis of the internal transcribed spacer region of the ribosomal RNA genes revealed approximately 80% similarity of *P. pachyrhizi* and *P. meibomiae* (Frederick et al. 2002). Of the 24 simple sequence repeat (SSR) markers developed for *P. pachyrhizi* only two of the markers yielded any detectable product when *P. meibomiae* genomic DNA was used as template. However the amount of product was negligible compared to the amount produced when *P. pachyrhizi* genomic DNA was used as template, which suggests that SSR loci either are lacking or the flanking DNA regions are not well conserved in *P. meibomiae* (Anderson et al. 2008).

Conserved gene sequences have been used to build phylogenetic trees, providing insight into fungal evolution and assisting in the clarification of confused phylogenies (Bullerwell et al. 2003; Seif et al. 2005). Mitochondrial (mt) genomes evolve independently of

Submitted 6 Aug 2009; accepted for publication 16 Dec 2009.

[1] Current address. Biophysics Laboratory, International Center of Physics-CIF, National University of Colombia, Ciudad Universitaria, Bogotá, Colombia

[2] Current address. Genome Project Solutions, 1024 Promenade Street, Hercules, California 94547

[3] Corresponding author. Email: Reid.Frederick@ars.usda.gov

the nuclear genomes and appear to evolve at an accelerated rate compared to that of the nuclear genome (Ballard and Whitlock 2004; Burger et al. 2003) and so often have been useful in cases where insufficient phylogenetic signal has accumulated in nuclear genes. While complete mt genomes are available for representative members of all fungal phyla, only seven members of the Basidiomycota have sequenced mt genomes. In this study the complete mtDNA sequences of *P. pachyrhizi* and *P. meibomiae* were determined and the genome organization, gene content and gene order were compared with other sequenced basidiomycetes. Analysis of the 14 common mt protein-coding genes was used to construct the phylogenetic relationships between these two *Phakopsora* species and other fungal mt genomes. These are the first descriptions of complete mt genome sequences for the rust fungi.

MATERIALS AND METHODS

Fungal isolates, library construction and sequence assembly.—*P. pachyrhizi* isolate Taiwan 72-1 and *P. meibomiae* isolate Puerto Rico were propagated on soybean and red kidney bean plants, respectively, as described by Frederick et al. (2002). Total DNA was isolated from urediniospores as described (Posada-Buitrago and Frederick 2005). Separate aliquots of total DNA preparations were shared randomly with a Hydroshear device (Genomics Solutions, Ann Arbor, Michigan) to generate fragments averaging either 3 kb or 8 kb. Fragments were cloned into pUC18 and pUC21, respectively, and sequenced with the Sanger sequencing method followed by capillary electrophoresis with ABI 3730 XL and MegaBase 4500 instruments following the whole genome shotgun (WGS) approach (Venter et al. 1998). Approximately 700 Mb of sequence was determined, the overwhelming amount of it being from the nuclear genome (toward goals independent of this report). Sequencing reads were assembled with the Phred-Phrap Package (Ewing and Green 1998, Ewing et al. 1998) and compared to known mt sequences in GenBank by BLAST analysis (Altschul et al. 1997) to identify the portion of the genome assembly corresponding to mtDNA. The complete mt sequences of *P. pachyrhizi* isolate Taiwan 72-1 and *P. meibomiae* isolate Puerto Rico have been submitted to GenBank as GQ332420 and GQ338834 respectively.

Sequence annotation.—Putative open reading frames were identified in both *Phakopsora* species with both ORF Finder at the NCBI and Artemis (Rutherford et al. 2000) using genetic code 4 (Fox 1987). Gene identities were assigned to ORF by screening for similarity with other organisms in the GenBank database with BLASTX (Gish and States 1993).

Introns were identified by BLASTX analysis and comparison with mtDNA of other fungi. An intron in the *cob* gene was found as described by Grasso et al. (2006). Exon and intron boundaries in the *cox1* gene were identified by comparison to the protein sequences of *Tilletia indica* (GenBank accession No. ABI95829), *Usitlago maydis* (GenBank accession No. YP_762688) and *Aspergillus niger* (GenBank accession No. YP_337885). The identification of introns was evaluated further by comparing mt sequences to the existing *P. pachyrhizi* germinating spore EST database (Posada-Buitrago and Frederick 2005).

tRNAs were identified with three independent programs, tRNAscan-SE (Lowe and Eddy 1997), DOGMA (Dual Organellar GenoMe Annotator) (Wyman et al. 2004) and ARAGORN (Laslett and Canback 2004). Codon usage was determined with the codon usage tool at http://www.ebioinfogen.com/biotools/codon-usage.htm with genetic code 4. Repeats were identified and analyzed with Tandem Repeats Finder (Benson 1999) and Palindrome and Einverted EMBOSS programs (Rice et al. 2000). Physical maps of the *Phakopsora* mtDNA were constructed with the GCView Server (Grant and Stothard 2008).

Comparative genomics.—The complete mt genomes of the seven fungi belonging to the *Basidiomycota* were retrieved from GenBank. Mitochondrial gene content and gene order of *P. pachyrhizi* and *P. meibomiae* were compared visually to these seven fungi.

Phylogenetic analysis.—The 14 protein-coding genes shared in common among the 58 fungal mt genomes were aligned with Clustal W (Thompson et al. 1994) with default parameters. The *rps3* gene was not included in this analysis because it is not conserved in some fungal mt genomes. Alignments were manually edited in SeaView (Galtier et al. 1996) to remove regions of uncertain alignment, and 3080 amino acid positions from the 14 concatenated protein sequences were used in the phylogenetic analysis. Maximum likelihood analysis was performed by PhyML (Guindon and Gascuel 2003) with the JTT substitution matrix for amino acid substitution (Jones et al. 1992) and gamma distribution. The value of the shape parameter was optimized by PhyML. Branch support was tested with 100 bootstrap replicates (Felsenstein 1985).

RESULTS

Gene content and genome organization.—The mt genomes of *P. pachyrhizi* and *P. meibomiae* are circular molecules of 31 825 bp and 32 520 bp respectively (FIG. 1). The mt genomes of both *Phakopsora* species contain 14 protein-coding genes, one ORF coding for the ribosomal protein (*rps3*), the large (*rnl*) and small (*rns*) ribosomal subunit RNAs, and 24 tRNAs. The protein-coding mt genes include three ATP synthase subunits (*atp*6, *atp*8 and *atp*9), three cytochrome oxidase subunits I, II and III (*cox*1, *cox*2, *cox*3), seven nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad*1–6, *nad*4L) and cytochrome b (*cob*) (TABLE I). In addition *P. pachyrhizi* and *P. meibomiae* also contain an ancient gene (*rps3*) coding for a ribosomal protein found in some Ascomycota, Basidiomycota, Zygomycota and Chyridiomycota (Bullerwell et al. 2000). All

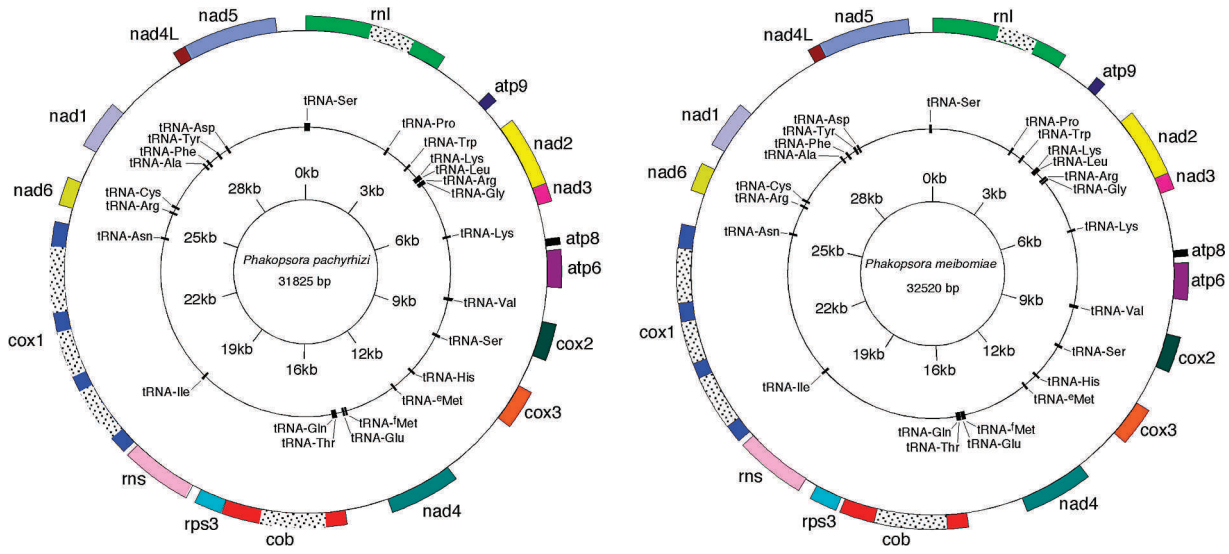

FIG. 1. Circular DNA maps of the mitochondrial genomes of *Phakopsora pachyrhizi*, left, and *P. meibomiae*, right. Genes for tRNA are located on the middle circle. The 14 protein-coding genes and the small and large ribosomal genes are depicted by shaded boxes on the outer circle. Introns are indicated by white boxes with black dots. All genes are on the same DNA strand and were transcribed clockwise.

mt genes are transcribed clockwise from the same strand, and the gene order is identical in *P. pachyrhizi* and *P. meibomiae* (FIG. 1). The percent similarity between the two *Phakopsora* species is 93% for the exons of the 14 protein-coding genes, 86% for the *rps3* ORF, and 94% and 91% for the ribosomal RNAs, *rnl* and *rns*, respectively. The percent similarity of the non-coding intergenic regions between *P. pachyrhizi* and *P. meibomiae* is 62%.

The mean GC content for the *P. pachyrhizi* and *P. meibomiae* mt genomes is 34.6% and 34.9%, respectively. Overall the *P. pachyrhizi* mt genome contains 35.4% A, 30% T, 21.8% G and 12.8% C, while the *P. meibomiae* mt genome is composed of 35.9% A, 29.2% T, 22.2% G and 12.7% C. Protein-coding and non-coding regions of the *P. pachyrhizi* mt genome have similar GC content with 36.2% and 31.5% respectively. The *P. meibomiae* mt genome has 34.9% and 32.5% GC content in the protein-coding versus non-coding regions, respectively.

Introns were identified in the *cox*1 *cob*, and *rnl* genes in both *Phakopsora* species. The *cox*1 gene of both species contains three introns. Introns 1 and 3 of the *cox*1 gene contain ORFs with amino acid similarity to the LAGLIDADG motif of endonucleases in other fungi, including the basidiomycete *Ustilago maydis* (GenBank accession No. YP_762694.1). Comparison of these two *cox*1 introns to the NCBI EST database provided evidence for their transcription, thereby supporting their identification as intronic ORFs. BLASTX analysis with intron 2 in the *cox1* gene did not reveal any significant amino acid similarity to any other organism, and no sequence similarity was found in comparisons to the *P. pachyrhizi* ESTs and the NCBI EST database. The intron in the *cob* gene was described in *P. pachyrhizi* (Grasso et al. 2006). Comparison of the *cob* intron and the flanking exons to the *P. pachyrhizi* ESTs revealed two clones with gaps in coverage corresponding to the intron, suggesting that this intron was not transcribed. However BLASTX analysis found similarity at the amino acid level to the LAGLIDADG endonucleases (RNA maturases) in other fungi, including the basidiomycetes *M. perniciosa* and *U. maydis* (Genbank accession Nos. YP_025851.1 and YP_762690.1 respectively).

Comparison of the *rnl* gene to *Phakopsora pachyrhizi* ESTs in GenBank suggests the presence of an intron. No *P. pachyrhizi* ESTs span the *P. pachyrhizi rnl* gene 1223–1924 bp or in the *P. meibomiae* mt 1253–1879 bp. Comparison of the *P. pachyrhizi* and *P. meibomiae rnl* genes to the ESTs of three other rust fungi, *Puccinia graminis* f. sp. *tritici* (GenBank accession No. CV191672), *P. striiformis* f. sp. *tritici* (GenBank ES321943) and *Uromyces appendiculatus* (GenBank EH299595), revealed a gap in coverage 1560–1924 bp in *P. pachyrhizi* and 1598–1879 bp in *P. meibomiae*, which indicates that these regions are introns. BLASTX analysis of these introns did not reveal any significant similarity to any other organisms in GenBank.

Both *P. pachyrhizi* and *P. meibomiae* mt genomes contain 10 perfect or near identical tandem repeats, ranging in size from 2–53 bp and with 2–32 copies (TABLE II). None of the tandem repeats are present in

TABLE I. Gene content of *Phakopsora pachyrhizi* and *P. meibomiae* mitochondrial genomes

Genetic element	*P. pachyrhizi* Location (nt)	*P. pachyrhizi* Codon Start	*P. pachyrhizi* Codon Stop	*P. meibomiae* Location (nt)	*P. meibomiae* Codon Start	*P. meibomiae* Codon Stop	Percent similarity[a]
rnl[b]	join: 2–1559; 1925–2905			join: 2–1597; 1880–2829			94.04
tRNA-Pro	2944–3015			2906–2977			
tRNA-Trp	3872–3943			3375–3446			
atp9	4028–4249	ATA	TAG	3603–3824	ATG	TAG	95.04
tRNA-Lys	4371–4442			4000–4071			
tRNA-Leu	4449–4531			4078–4160			
tRNA-Arg	4585–4655			4441–4511			
tRNA-Gly	4660–4731			4527–4598			
nad2	4745–6163	ATG	TAA	4612–6030	ATG	TAA	91.97
nad3	6165–6536	ATG	TAA	6032–06379	ATG	TAA	95.69
tRNA-Lys	6702–6773			6538–6609			
atp8	7268–7414	ATA	TAA	7630–7776	ATT	TAA	96.60
atp6	7504–8271	ATG	TAA	7900–8667	ATG	TAA	94.40
tRNA-Val	8878–8948			9270–9340			
cox2	9000–9755	ATG	TAA	9427–10182	ATA	TAA	93.65
tRNA-Ser	10211–10297			10780–10866			
cox3	10441–11247	ATG	TAA	11057–11860	ATA	TAA	95.15
tRNA-His	11745–11817			12114–12186			
tRNA-[e]Met[c]	12582–12652			12663–12733			
nad4	12732–14165	ATG	TAG	12870–14303	ATG	TAG	90.80
tRNA-[f]Met[d]	14464–14535			15095–15166			
tRNA-Glu	14554–14626			15168–15240			
tRNA-Thr	14850–14923			15264–15337			
tRNA-Gln	14932–15003			15342–15413			
cob[e]	join: 15086–15514; 16851–17609	ATG	TAG	join: 15531–15959; 17452–18183	ATG	TAA	92.94
rps3	17626–18186	ATG	TAA	18268–18843	ATG	TAG	86.07
rns	18308–19732			18983–20423			91.55
tRNA-Ile	19775–19847			20541–20613			
cox1[f]	join: 19904–20287; 21407–21730; 22665–23054; 24385–24876	ATA	TAA	join: 20692–21075; 22205–22528; 23397–23786; 24910–25401	ATA	TAA	96.67
tRNA-Asn	25006–25077			25773–25844			
nad6	25217–25786	ATG	TAA	26113–26697	ATG	TAG	95.23
tRNA-Arg	25938–26008			26832–26902			
tRNA-Cys	26079–26150			26975–27046			
nad1	26488–27486	ATG	TAA	27134–28126	ATG	TAG	95.26
tRNA-Ala	27940–28010			29031–29101			
tRNA-Phe	28107–28179			29276–29348			
tRNA-Tyr	28541–28621			29595–29675			
tRNA-Asp	28917–28988			29733–29804			
nad4L	29068–29328	ATG	TAG	29886–30146	ATG	TAG	94.64
nad5	29331–31220	ATG	TAA	30149–32029	ATG	TAA	92.22
tRNA-Ser	31964–31778			32399–32483			

[a] Percent similarity calculated between exons of *P. pachyrhizi* and *P. meibomiae*.
[b] The *rnl* gene contains one intron.
[c] Methionine tRNA translation elongation tRNA.
[d] N-formyl-methionyl-tRNA translation initiation tRNA.
[e] The *cob* gene contains one intron.
[f] The *cox1* gene contains three introns.

both *P. pachyrhizi* and *P. meibomiae* mt genomes. In *P. pachyrhizi* the dyad TA is the most repeated motif with 32 copies, followed by the 5 bp perfect tandem repeats CTTCT and AGATA with 18.4 and 11.8 copies respectively. The largest tandem repeat motif in the *P. pachyrhizi* mt genome is 28 bp and occurs three times after *cox2* (position 9842–9925). In *P. meibomiae* the most repeated tandem motifs are GATATA and ATATCA with 11 and 15.5 copies, respectively. These two tandem near perfect repeat sequences also comprise part of a 68 bp inverted repeat element at positions 24730–24797 and 25630–25697, which flank exon 4 of *cox1*. The largest tandem repeat unit in *P. meibomiae* is 53 bp and occurs at 2.0 copies after *nad1*. Four inverted repeat motifs were found in the *P. pachyrhizi* mt DNA, whereas two inverted repeats were identified in *P. meibomiae* (TABLE II).

Codon usage and tRNA genes.—The four most frequent amino acids in the *P. pachchyrhizi* mt genome are isoleucine (593), leucine (478), serine (423) and valine (400), and these same four amino acids are the most common in the *P. meibomiae* mt genome with 594, 477, 417 and 385 counts respectively (TABLE III). The five most frequent codons in the mt genomes of *P. pachyrhizi* and *P. meibomiae* are CTA, ATA, GTA, TTT and ATT accounting for 34.9% and 34.7% of all codons in *P. pachyrhizi* and *P. meibomiae*, respectively. The codons CGC (Arg) and TTG (Leu) are not used in the mt genome of either *Phakopsora* species, and codons GGC (Gly) and TTA (Leu) are not used in *P. pachyrhizi* and *P. meibomiae*, respectively. Fourteen codons were underrepresented in both *Phakopsora* species, occurring 1–15 times each (TABLE III). A codon usage bias exists toward codons with A or T in the third position. Codons ending in A or T comprise 74.6% and 75.2% of all codons of *P. pachyrhizi* and *P. meibomiae* respectively.

Most protein-coding genes start with the canonical translation initiation codon ATG. However *atp*9, *atp*8 and *cox*1 genes of *P. pachyrhizi* and *cox*1, *cox*2, and *cox*3 genes of *P. meibomiae* appear to use the ATA start codon while *atp*8 of *P. meibomiae* displays a potential ATT start codon. In both *Phakopsora* species the preferred stop codon in the mt is TAA, occurring in 11 genes of *P. pachyrhizi* and nine genes of *P. meibomiae*. The alternative stop codon TAG occurs in four mt genes of *P. pachyrhizi* and six mt genes of *P. meibomiae*, while the universal stop codon TGA was not observed but instead encodes for tryptophan as has been described for other fungal mt genomes (Saccone et al. 2002).

Twenty-four tRNAs were identified in each of *P. pachyrhizi* and *P. meibomiae* mt genomes (TABLE I). tRNA genes are located on the same DNA strand as the other genes, and the order is identical in both *Phakopsora* species (FIG. 1). Twelve of the tRNA group into three clusters of four tRNA genes each, while the other 12 tRNA occur singly (FIG. 1).

Comparative genomics.—Comparison of the mt genomes of *P. pachyrhizi* and *P. meibomiae* with those from other *Basidiomycota* did not show overall synteny of the protein coding genes (FIG. 2). However closer inspection reveals trends in the grouping of some genes. In eight of the nine species examined *nad*2 and *nad*3 cluster together on the genome as do *nad*4L and *nad*5. The linkage of these genes is such that no intergenic spacers exist between *nad*2 and *nad*3 or between *nad*4L and *nad*5 in *P. pachyrhizi* and *P. meibomiae*. Similar linkage between *nad* genes had been reported among the *Basidiomycota* (Wang et al. 2008; Forget et al. 2002). In addition the *atp*6 and *atp*8 genes group together in the mt genomes of *P. pachyrhizi*, *P. meibomiae*, *T. indica*, *T. walkeri* and *U. maydis*. In some other cases this is due to genes being translated from a bicistronic transcript, but this remains to be investigated for these organisms.

P. pachyrhizi and *P. meibomiae* each contain three introns in *cox*1 and one intron in *cob*. Likewise other members of the *Basidiomycota* show a bias for insertion of introns in these two protein-coding genes. Highly conserved sequence regions are the preferred targets of introns. *Cox*1 and *cob* are the most conserved mt genes and therefore the most frequent targets for insertion elements in fungal mt genomes (Paquin et al. 1997).

Phylogenetic analysis.—Phylogentic trees were built with the 14 protein-coding genes (*atp*6, *atp*8, *atp*9, *cox*1, *cox*2, *cox*3, *cob*, *nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5 and *nad*6) in common from 58 fungal species. The phylogenetic trees obtained for each of the mt genes showed only minor incongruence with commonly accepted fungal taxonomy (data not shown). Sequences of the 14 protein-coding mt genes were concatenated and re-analyzed to produce a final phylogenetic tree (FIG. 3). The topology of this tree showed higher bootstrap (BS) support for most of the branches than those built with independent genes and agreed with commonly accepted fungal taxonomy and recent fungal phylogenies (Kouvelis et al. 2004, Blackwell et al. 2006, Formighieri et al. 2008).

DISCUSSION

The mt genomes of fungi show tremendous size variation ranging from the smallest known at 11 094 bp in *Hanseniaspora uvarum* (Pramateftaki et al. 2006) to the largest at 109 103 bp in *Moniliophthora perniciosa* (Formighieri et al. 2008). With the

TABLE II. Repeat sequences in *Phakopsora pachyrhizi* and *P. meibomiae* mitochondrial genomes

Tandem repeats

P. pachyrhizi

Location	Period size	Consensus pattern	Copy
3141–3202	2	TA	32.0
3338–3416	24	ATATATATACCAATATAGATACAG	3.3
8344–8392	18	AATACTAAAACTAAAACT	2.7
8389–8442	19	ACTATGTAAACTTAACAAT	2.8
9842–9925	28	TGTGTAAATGTATAATATCTGGAGTATA	3.0
11315–11366	15	TTATCTTACCTTATC	3.5
12065–12104	20	AGGTATATACAATACATTAT	2.0
19739–19766	7	AGAAGAT	4.0
27593–27651	5	AGATA	11.8
31226–31317	5	CTTCT	18.4

P. meibomiae

6819–6870	5	CTAAG	10.4
7147–7200	6	TAAGTA	9.0
20484–20514	15	AATATGATGAGGGTA	2.0
24733–24798	6	GATATA	11.0
25625–25716	6	ATATCA	15.5
25931–26002	11	AGAATATAATA	6.3
28170–28221	23	GATAAAATGAAAATAATAAGTAC	2.3
28426–28531	53	TATACCCTAAGGATAATATACCATTATGAGGATATACCCGATCATTATGAGGA	2.0
28596–28635	18	ATTATGATCAGGTACACA	2.2
32265–32354	21	TAACAACAGCATCAATAACGC	4.3

Palindromes

P. pachyrhizi

Location	Size	Sequence
3157–3176	20	TATATATATATTATATATAT
3202–3183		ATATATATATAATATATATA
11959–11968	10	ATATATAATA
11980–11971		TATATATTAT
24072–24081	10	ATATCATATA
24091–24082		TATAGTATAT

P. meibomiae

8375–8384	10	TACTAGTACC
8394–8385		ATGATCATGG
14483–14492	10	ATATATATAT
14502–14493		TATATATATA
25565–25574	10	ATATATATAT
25584–25575		TATATATATA

Inverted repeats

P. pachyrhizi

Location	Matches	Sequence
2796–2836	33/41	ATATTTATATAT–GTATATATATATTAAAAAATTAATATAGG
3180–3139		TATATATATATATTATATATATATATATATTATATGATATCC
6995–7048	42/53	ATATGTGATATATATA–GATATAGGATATACAGTATGTAATATATGATATATATA
7130–7077		TATATAAGATATATATAGTATATAATATATGACATATAATATATA–TGGATATAT
9963–10020	45/56	ATATAAAAGAAGGAGTAGAGA–TATGTAAAGAATATATGGGGTATATATAATATATGGA

TABLE II. Continued

Inverted repeats		
P. pachyrhizi		
Location	Matches	Sequence
11393–11337		TATATTCTCTT–CTAATCCCTAATATATTTCTTTTCTAATCCAT–TCTATTCTATTCCT
24063–24117	44/55	TAGATATAAATATCATATATATAT–GATATATCTAT–TAATATAGATATTAATATAT
24180–24124		ATCTATAGATAGAGTATAGAAATAGATATATAGATATAGTATATGGATATATATATA
P. meibomiae		
6474–6497	21/24	ATATATATATATATAGATATGATA
7294–7271		TATATATATATATATAAATAATAT
24730–24797	50/68	ATGGATATAGGTATAGATATAAATATAGATACAGATATAAATATTAATATAGATATAGGTATAAATAT
25697–25630		TAACTATAGCTATGACTATAACTATAGCTATGACTATAACTATAACTATGACTATAACTATAACTATA

TABLE III. Codon usage in protein-coding and *rns* genes of *Phakopsora pachyrhizi* and *P. meibomiae* mitochondrial genomes

Codon	AA[a]	Ppa[b]	Pme[c]	Codon	AA	Ppa	Pme	Codon	AA	Ppa	Pme
GCG	Ala	95	90	**AAA**[d]	Lys	9	10	AGT	Ser	175	168
GCA	Ala	112	105	TTG	Leu	0	0	**AGC**	Ser	19	21
GCT	Ala	44	48	TTA	Leu	2	0	TCG	Ser	38	38
GCC	Ala	14	25	CTG	Leu	52	61	**TCA**	Ser	145	134
TGT	Cys	40	43	**CTA**	Leu	388	384	TCT	Ser	36	47
TGC	Cys	2	1	CTT	Leu	34	29	TCC	Ser	10	9
GAT	Asp	101	99	CTC	Leu	2	3	ACG	Thr	77	64
GAC	Asp	10	12	**ATG**	Met	94	94	**ACA**	Thr	140	144
GAG	Glu	89	79	AAT	Asn	102	113	ACT	Thr	33	46
GAA	Glu	26	38	**AAC**	Asn	33	27	ACC	Thr	9	10
TTT	Phe	240	245	CCG	Pro	47	35	GTG	Val	49	47
TTC	Phe	45	47	CCA	Pro	65	64	**GTA**	Val	321	309
GGG	Gly	115	107	CCT	Pro	29	37	GTT	Val	28	28
GGA	Gly	74	70	**CCC**	Pro	6	9	GTC	Val	2	1
GGT	Gly	107	101	CAG	Gln	53	49	TGG	Trp	20	23
GGC	Gly	0	2	**CAA**	Gln	38	44	**TGA**	Trp	52	50
CAT	His	77	77	AGG	Arg	59	56	TAT	Tyr	180	182
CAC	His	14	15	**AGA**	Arg	39	40	**TAC**	Tyr	38	30
ATA	Ile	326	319	CGG	Arg	3	5	TAG	*[e]	4	6
ATT	Ile	234	240	**CGA**	Arg	7	4	TAA	*	11	9
ATC	Ile	33	35	CGT	Arg	10	12				
AAG	Lys	66	69	CGC	Arg	0	0				

[a] Amino acid.
[b] *Phakopsora pachyrhizi.*
[c] *Phakopsora meibomiae.*
[d] For the codons in bold letters, the corresponding mitochondrial encoded *trn* genes have been identified.
[e] Stop codon.

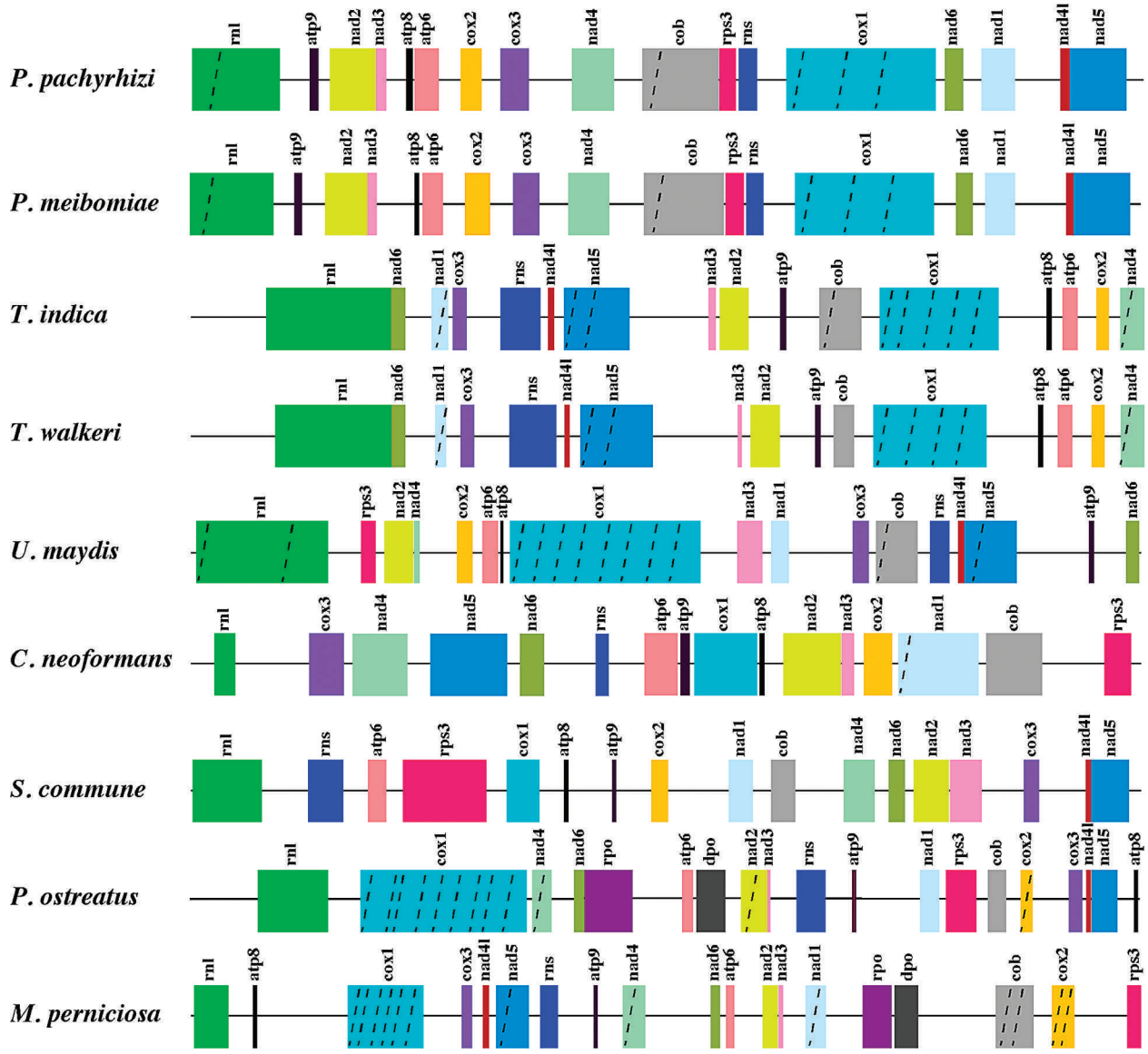

FIG. 2. Comparison of the gene content and order of the known mitochondrial genomes of *Phakopsora pachyrhizi* and *P. meibomiae* and other members of the *Basidiomycota*: *Tilletia indica* (DQ993184), *Tilletia walkeri* (NC_010651), *Ustilago maydis* (NC_008368), *Cryptococcus neoformans* var. *grubii* (NC_004336), *Schizophyllum commune* (NC_003049), *Pleurotus ostreatus* (NC_009905) and *Moniliophthora perniciosa* (AY376688). Gene sizes and total mitochondrial genome lengths are drawn proportionally. Protein coding and rRNA genes are represented by boxes. Dashed lines within genes represent the presence of introns.

sequences of the *P. pachyrhizi* and *P. meibomiae* mt genomes the number of *Basidiomycota* mt genomes available is nine. They were the first mt genomes from class *Pucciniomycetes* or from any rust or obligate fungal plant pathogen to be sequenced. The mt genomes of *P. pachyrhizi* and *P. meibomiae*, at 31 825 bp and 32 529 bp respectively, are of slightly less than average size reported for fungi. Among the *Basidiomycota* only *Cryptococcus neoformans* showed a smaller mt genome than these two *Phakopsora* species. The five other *Basidiomycota* mt genomes exceed 49 kbp and include the largest fungal mt genome sequenced to date, *M. perniciosa* (Formighieri et al. 2008). Despite their relative small size among the *Basidiomycota*, both *Phakopsora* species contain the full complement of 14 protein coding genes and two ribosomal nuclear genes as is typical of fungi, along with 24 tRNAs. In addition *P. pachyrhizi* and *P. meibomiae* contain *rps3* like five other members of the Basidiomycota, with the two *Tilletia* species being the exceptions.

While the mt genomes of some groups of fungi, particularly the *Sordariomycetes* and *Eurotiomycetes* classes of *Ascomycota*, show a high degree of synteny

FIG. 3. The single phylogenetic tree constructed from unambiguously aligned portions of concatenated protein-coding sequences of all 14 essential mitochondrial genes. The topology shown was inferred with PhyML and the JTT model of protein evolution. Maximum likelihood-bootstrap support was calculated from 100 replicates.

(Kouvelis et al. 2004; Cardoso et al. 2007), this was not evident among the *Basidiomycota*. With the addition of two *Phakopsora* species five different classes of the *Basidiomycota* have been sequenced. Even at the family level the three sequenced members of the *Agaricales* do not show any significant synteny. However the two *Phakopsora* species included in this study possess the same gene order and structure, including the presence of introns. Similarly *T. indica* possess the same gene order as *T. walkeri*, and *M. perniciosa* and *M. roreri* exhibited a high degree of synteny, so there is a conservation of gene order at least for these congeners. This data suggests that gene order among the *Basidiomycota* is re-arranged during speciation (Formighieri et al. 2008). The mt gene order is consistent at the genus level but appears to be absent at the family level. However the *Agaricales* is the only family in which there are mt genome sequences available from multiple genera. The order of *cox2*, *nad1* and *cob* is conserved in *C. neoformans* and *S. commune* even though they are in different families. As more mt genomes for other members of the *Basidiomycota* become available a more comprehensive evolutionary history of gene order re-arrangement will unfold.

Despite the lack of synteny in mt gene order among the *Basidiomycota*, the protein-coding genes are consistent in their size. While *M. perniciosa* has the largest and *C. neoformans* the smallest mt genome among *Basdisiomycota*, there is less than a 300 bp overall difference between the protein-coding regions of these two species. The greatest variation between protein-coding regions within the *Basidiomycota* is just over 2100 bp, which occurs between *S. commune* and *P. meibomiae*. The variation in mt genome size is the result of differences in size and organization of intergenic spacers, number and size of optional introns and undetermined ORFs. The tremendous mt size variation between two strains of *C. neoformans* is attributable to differences in the number of introns contained within several of the mt genes (Litter et al. 2005). *P. anserina* is known to have larger sizes and numbers of introns, while *M. perniciosa* has larger intergenic regions between known mt genes (Wang et al. 2008; Formighieri et al. 2008). Because *S. commune* does not contain any

introns in the mt protein-coding genes the size variation compared to other *Basidiomycotina* mt genomes must stem from variations in the intergenic spacer regions (Forget et al. 2002).

Like most other fungal mt genomes, the *Phakopsora* mt genomes show a strong A + T bias. In addition a codon usage bias exists with an A or T in the third position in 74.6% and 75.2% of all codons in *P. pachyrhizi* and *P. meibomiae*, respectively. This codon bias most likely results from the strong mutational bias toward A and T of the mt genome (Xia 1996).

Analysis of the *P. pachyrhizi* and *P. meibomiae* mt genomes has identified potentially useful regions for the evaluation of inter- and intraspecies variation for population studies. While studies of fungal mt genomes have identified variation in the size and number of introns as a tool for distinguishing between strains of a single species (Litter et al. 2005), this approach is not likely to be useful for *Phakopsora* species because *P. pachyrhizi* and *P. meibomiae* possess identical introns in their mt genomes. The minor sequence variation within the protein-coding regions of the mt of *P. pachyrhizi* and *P. meibomiae* indicates that these might not be useful regions for examination of intraspecies variation. However the mt intergenic regions show only 62% similarity between *P. pachyrhizi* and *P. meibomiae*, so it might be plausible to develop intraspecies-specific molecular markers from these regions, as was accomplished for *Metarhizium anisopliae var. anisopliae* (Ghikas et al. 2006).

ACKNOWLEDGMENTS

We thank Paul Tooley and Doug Luster for their critical review of this manuscript. The use of trade, firm or corporation names in this publication is for the information and benefit of the reader. Such use does not constitute an official endorsement or approval by the USDA Agricultural Research Service of any product or service to the exclusion of others that may be suitable. This work was performed partly under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory, under contract No. DE-AC02-05CH11231.

LITERATURE CITED

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipaman TJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

Anderson SJ, Stone CL, Posada-Buitrago ML, Boore JL, Neelam BA, Stephens RM, Luster DG, Frederick RD, Pedley KF. 2008. Development of simple sequence repeat markers for the soybean rust fungus *Phakopsora pachyrhizi*. Mol Ecol Res 8:1310–1312.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Mol Ecol 13:729–744.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580.

Blackwell M, Hibbett DS, Taylor JW, Spatafora JW. 2006. Research coordination networks: a phylogeny for kingdom Fungi (Deep Hypha). Mycologia 98:829–837.

Bonde MR, Peterson GL, Dowler WM. 1988. A comparison of isozymes of *Phakopsora pachyrhizi* from the eastern hemisphere and the New World. Phythopathology 78: 1491–1494.

Bromfield KR. 1984. Soybean rust. Monograph No. 11. St Paul: Am Phytopathol Soc.

Bullerwell CE, Burger G, Lang BF. 2000. A novel motif for identifying *rps3* homologs in fungal mitochondrial genomes. Trends Biochem Sci 25:363–365.

———, Forget L, Lang BF. 2003. Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. Nucleic Acids Res 31:1614–1623.

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: Anything goes. Trends Genet 19:709–716.

Cardoso MA, Tambor JH, Nobrega FG. 2007. The mitochondrial genome from the thermal dimorphic fungus *Paracoccidioides brasiliensis*. Yeast 24:607–616.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred II. Error probabilities. Genome Res 8:186–194.

———, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred I. Accuracy assessment. Genome Res 8:175–185.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Forget L, Ustinova J, Wang Z, Huss VA, Lang FB. 2002. *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing and an evolutionary link to lower fungi. Mol Biol Evol 19:310–319.

Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, Góes-Neto A, Cotomacci C, Carazzolle MF, Sardinha-Pinto N, Thomazella DPT, Rincones J, Digiampietri L, Carraro DM, Azeredo-Espin AM, Reis SF, Deckmann AC, Gramacho K, Gonçalves MS, Moura Neto JP, Barbosa LV, Meinhardt LW, Cascardo JCM, Pereira GAG. 2008. The mitochondrial genome of the phytopathogenic basidiomycete *Moniliophthora perniciosa* is 109 kb in size and contains a stable integrated plasmid. Mycol Res 112:1136–1152.

Fox TD. 1987. Natural variation in the genetic code. Ann Rev Genet 21:67–91.

Frederick RD, Snyder CL, Peterson GL, Bonde MR. 2002. Polymerase chain reaction assays for the detection and discrimination of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomiae*. Phytopathology 92: 217–227.

Galtier N, Gouy M, Gautier C. 1996. SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548.

Ghikas DV, Kouvelis VN, Typas MA. 2006. The complete mitochondrial genome of the entomopathogenic

fungus *Metarhizium anisopliae* var. *anisopliae*: gene order and *trn* gene clusters reveal a common evolutionary course for all Sordariomycetes, while intergenic regions show variation. Arch Microbiol 185:393–401.

Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. Nat Genet 3:266–272.

Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184.

Grasso V, Palermo S, Sierotzki H, Garibaldi A, Gisi U. 2006. Cytochrome b gene structure and consequences for resistance to Qo inhibitor fungicides in plant pathogens. Pest Manage Sci 62:465–472.

Guindon S, Gascuel O. 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704.

Hennings P. 1903. Some new Japanese Uredinales. Hedwiga 4(suppl.):107–108.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282.

Kouvelis VN, Ghikas DV, Typas MA. 2004. The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanii*) suggests a minimum common gene organization in mtDNA of Sordariomycetes: phylogenetic implications. Fungal Genet Biol 41:930–940.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16.

Levy C. 2005. Epidemiology and chemical control of soybean rust in southern Africa. Plant Dis 89:669–674.

Litter J, Keszthelyi A, Hamari Z, Pfeiffer I, Kucsera J. 2005. Differences in mitochondrial genome organization of *Cryptococcus neoformans* strains. Antonie van Leeuwenhoek 88:249–255.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.

Miles MR, Frederick RD, Hartman GL. 2003. Soybean rust: Is the U.S. soybean crop at risk?. St Paul: Am Phytopathol Soc, Online publication http://www.apsnet.org/online/feature/rust/

Ono Y, Buritica P, Hennen JF. 1992. Delimination of *Phakopsora*, *Physopella* and *Cerotelium* and their species of Leguminosae. Mycol Res 96:825–850.

Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF. 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Curr Genet 31: 380–395.

Posada-Buitrago ML, Frederick RD. 2005. Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. Fungal Genet Biol 42:949–962.

Pramateftaki PV, Kouvelis VN, Lanaridis P, Typas MA. 2006. The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. FEMS Yeast Res 6: 77–90.

Rice P, Lungden I, Bleasby A. 2000. EMBOSS: The European molecular biology open software suite. Trends Genet 16:276–277.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944–945.

Saccone C, Gissi C, Reyes A, Larizza A, Sbisà E, Pesole G. 2002. Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene 286:3–12.

Schneider RW, Hollier CA, Whitham HK, Palm ME, McKemy JM, Hernandez JR, Levy L, Devries-Paterson R. 2005. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. Plant Dis 89:774.

Seif E, Leigh J, Liu Y, Roewer I, Forget L, Lang BF. 2005. Comparative mitochondrial genomics in zygomycetes: bacteria-like Rnase P RNA, mobile elements and a close source of the group I intron invasion in angiosperms. Nucleic Acids Res 33:734–744.

Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.

Vakili NG, Bromfield KR. 1976. *Phakopsora* rust on soybean and other legumes in Puerto Rico. Plant Dis Rep 60: 995–999.

Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. 1998. Shotgun sequencing of the human genome. Science 280:1540–1542.

Wang Y, Zeng F, Hon CC, Zhang Y, Leung FCC. 2008. The mitochondrial genome of the Basidiomycete fungus *Pleurotus ostreatus* (oyster mushroom). FEMS Microbiol Lett 280:34–41.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.

Xia X. 1996. Maximizing transcription efficiency causes codon usage bias. Genetics 144:1309–1320.

Yorinori JT, Paiva WM, Frederick RD, Costamilan LM, Bertagnoli PF, Hartman GL, Godoy CV, NunesJrJ. 2005. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. Plant Dis 29:675–677.